Determining Possible Fragments Based On Experimental Restrictions

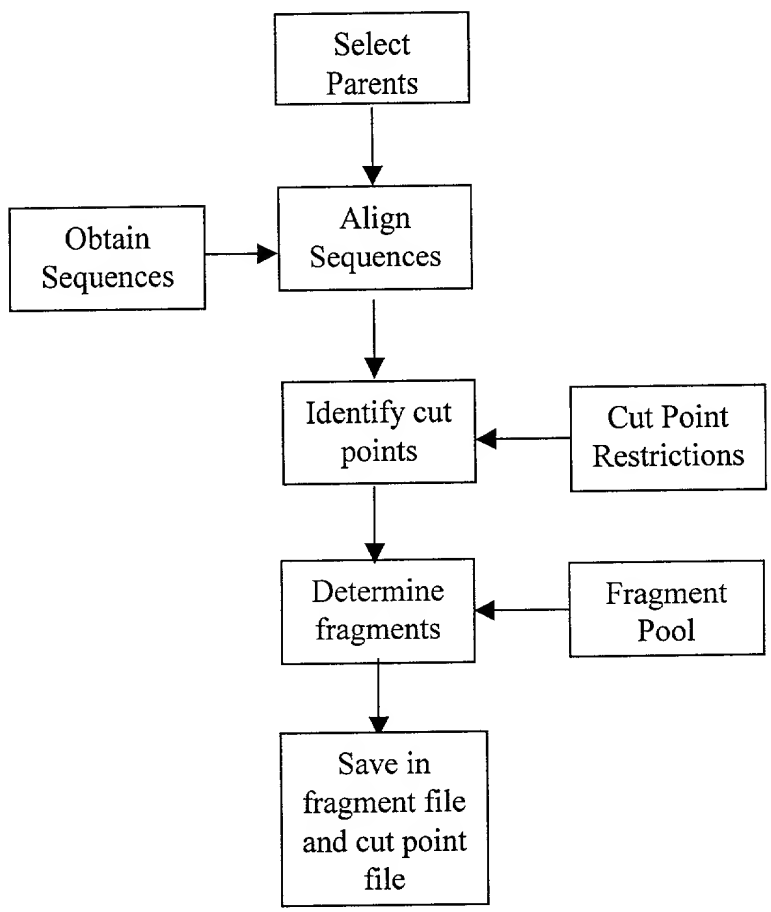


FIG. 1A

Determining the Schema Disruption Profile for a Structure

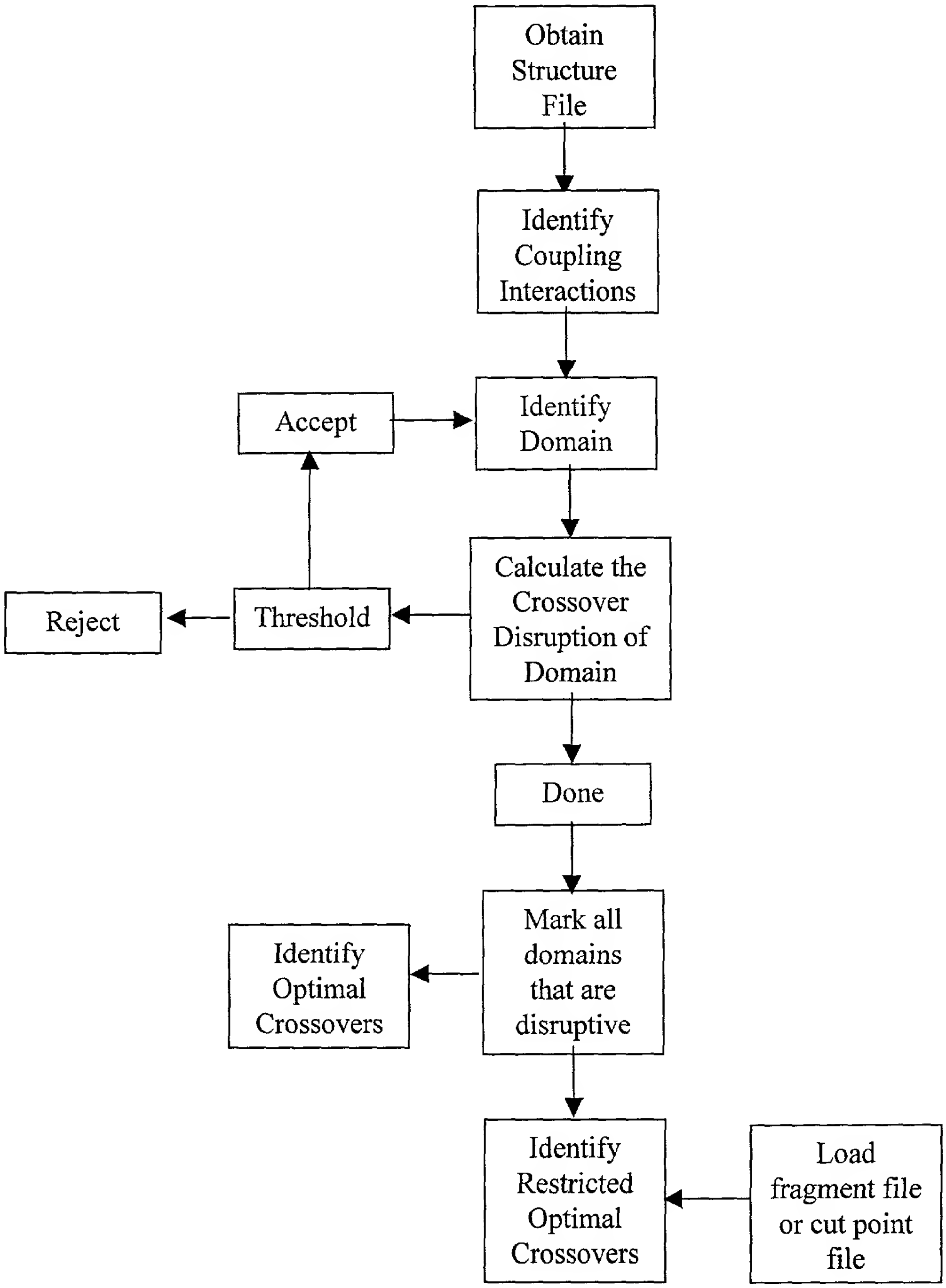


FIG. 1B

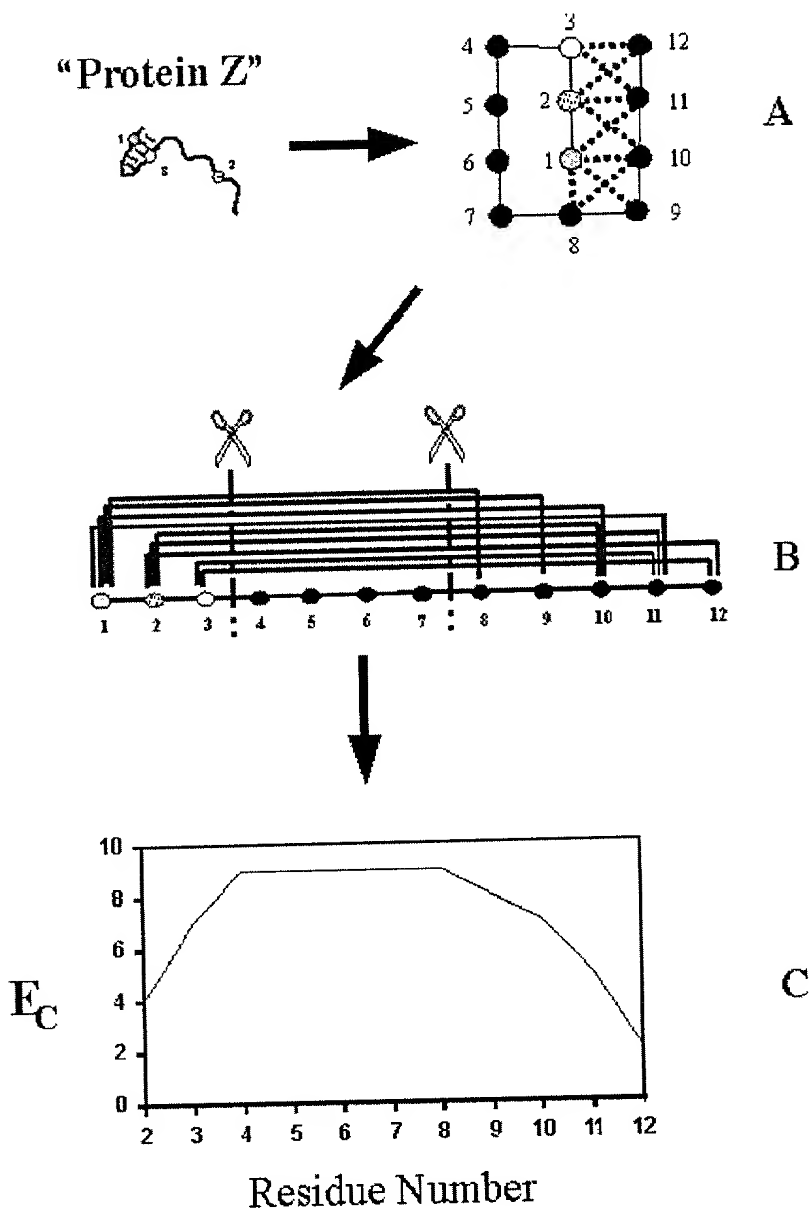


FIG. 2

```
1     1   TPVSEKQLAEVVANTITPLMKAQSVPGMAVAVIYQGKPHYYTFGKADIAA
2     1   AAKTEQQIADIVNRTITPLMQEQAIPGMAVAIIYEGKPYYFTWGKADIAN
3     1   TKLTELQVATIVNNTLTPLLEKQGIPGMAVAVFYDGKPQFFNYGMADIKA
4     1   YARGEAPLTAAVDGIIQPMLKEYRIPGMAVAVLKDGKAHYFNYGVANRES

1    51   NKPVTPQTLFELGSISKTFTGVLGGDAIARGEISLDDAVTRYWPQLTGKQ
2    51   NHPVTQQTLFELGSVSKTFNGVLGGDRIARGEIKLSDPVTKYWPELTGKQ
3    51   GRPVTENTLFELGSVSKTFTGVAGEYAMQTGIMNLNDPVTEYAPELTGSQ
4    51   GQRVSEQTLFEIGSVSKTLTATLGAYAAVKGGFELDDKVSQHAPWLKGSA

1   101   WQGIRMLDLATYTAGGLPLQVPDEVTDNASLLRFYQNWQPQWKPGTTRLY
2   101   WRGISLLHLATYTAGGLPLQIPGDVTDKAELLRFYQNWQPQWTPGAKRLY
3   101   WKDVKMLHLATYTAGGLPLQLPDSVTDQKSLWQYYQQWQPQWAPGVMRNY
4   101   FDGVTMAELATYSAGGLPLQFPDEVDSNDKMRTYYRHWSPVYPAGTHRQY

1   151   ANASIGLFGALAVKPSGMPYEQAMTTRVLKPLKLDHTWINVPKAEEAHYA
2   151   ANSSIGLFGALAVKSSGMSYEEAMTRRVLQPLKLAHTWITVPQSEQKNYA
3   151   SNASIGLFGALAVKRSQLTFENYMKEYVFQPLKLDHTFITIPESMQSNYA
4   151   SNPSIGLFGHLAANSLGQPFEQLMSQTLLPKLGLHHTYIQVPESAIANYA

1   201   WGYRDGKAVRVSPGMLDAQAYGVKTNVQDMANWVMANMAPENVADASLKQ
2   201   WGYLEGKPVHVSPGQLDAEAYGVKSSVIDMARWVQANMDASHVQEKTLQQ
3   201   WGYKDGQPVRVTLGMLGEEAYGVKSTSQDMVRFMQANMDPESLGNDKLKE
4   201   YGYKEDKPVRVTPGVLAAEAYGIKTGSADLLKFTEANMGYQ. GDAALKT

1   251   GIALAQSRYWRIGSMYQGLGWEMLNWPVEANTVVEGSDSKVALAPLPVAE
2   251   GIELAQSRYWRIGDMYQGLGWEMLNWPLKADSIINGSDSKVALAALPAVE
3   251   AIIASQSRYFQAGDMFQGLGWEMYSWPINPQGVIADSGNDIALKPRKVEA
4   249   RIALTHTGFYSVGDMTQGLGWESYAYPLTEQALLAGNSPAVSFQANPVTR

1   301   VNPPAPPVKASWVHKTGSTGGFGSYVAFIPEKQIGIVMLANTSYPNPARV
2   301   VNPPAPAVKASWVHKTGSTGGFGSYVAFVPEKNLGIVMLANKSYPNPARV
3   301   LVPAQPAVRASWVHKTGATNGFGAYIVFIPEEKVGIVMLANKNYPNPVRV
4   299   FAVPKAMGEQRLYNKTGSTGGFGAYVAFVPARGIAIVMLANRNYPIEARV

1   351   EAAYHILEALQ
2   351   KAAWRILEKLQ
3   351   QAAYDILQALR
4   349   KAAHAILSQLA
```

1 Enterbacter cloacae P05364 (X03866)
2 Citrobacter freundii P05193 (X07274)
3 Yersinia enterocolitica P45460 (X63149)
4 Klebsiella pneumoniae Q48437 (X77455)

FIG. 3

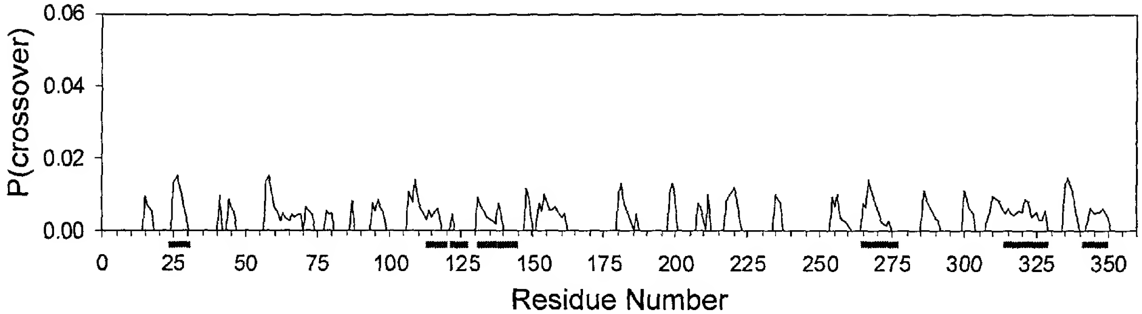

P(crossover)
0.06
0.04
0.02
0.00
0
25
50
75
100
125
150
175
200
225
250
275
300
325
350
Residue Number

FIG. 4A

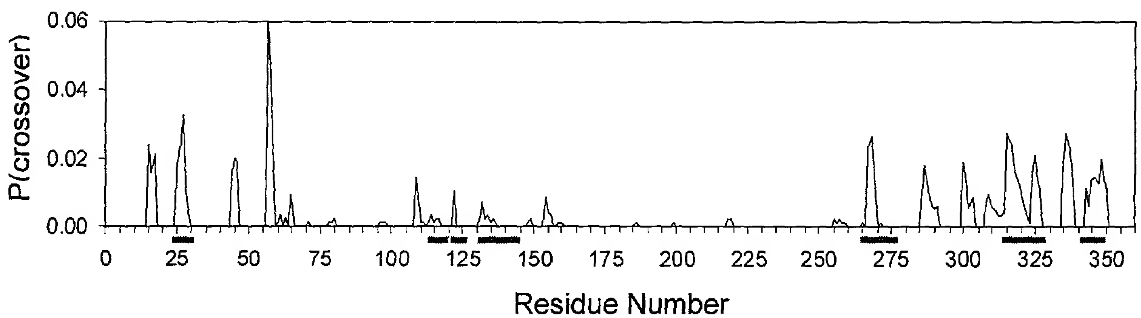

P(crossover)
0.06
0.04
0.02
0.00
0
25
50
75
100
125
150
175
200
225
250
275
300
325
350
Residue Number

FIG. 4B

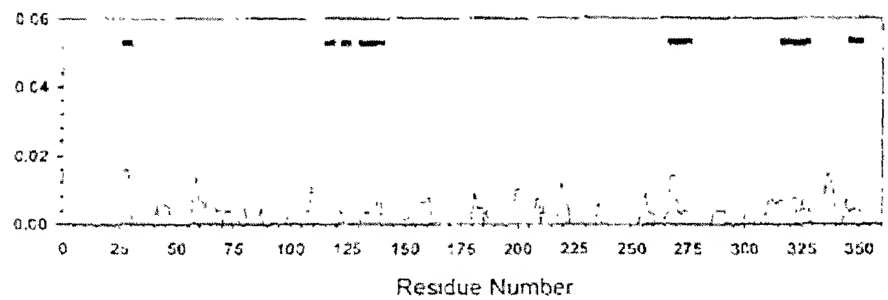

0.06
0.04
0.02
0.00
0
25
50
75
100
125
150
175
200
225
250
275
300
325
350
Residue Number

FIG. 4C

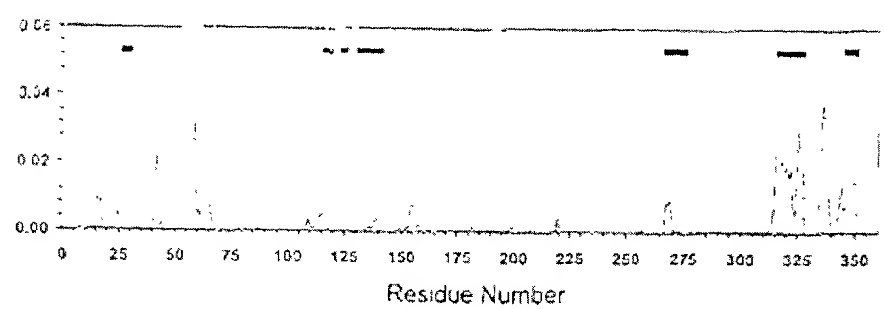
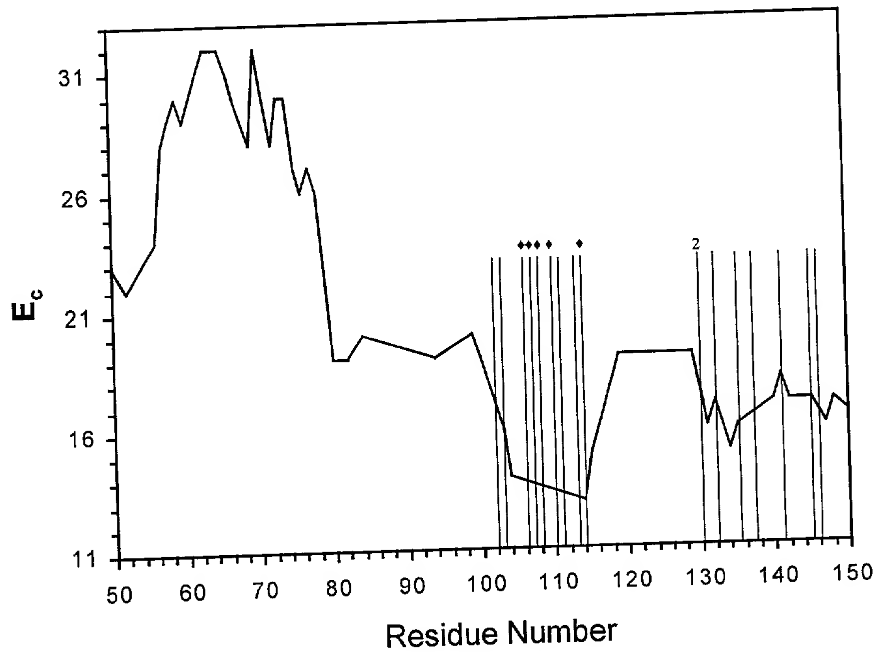
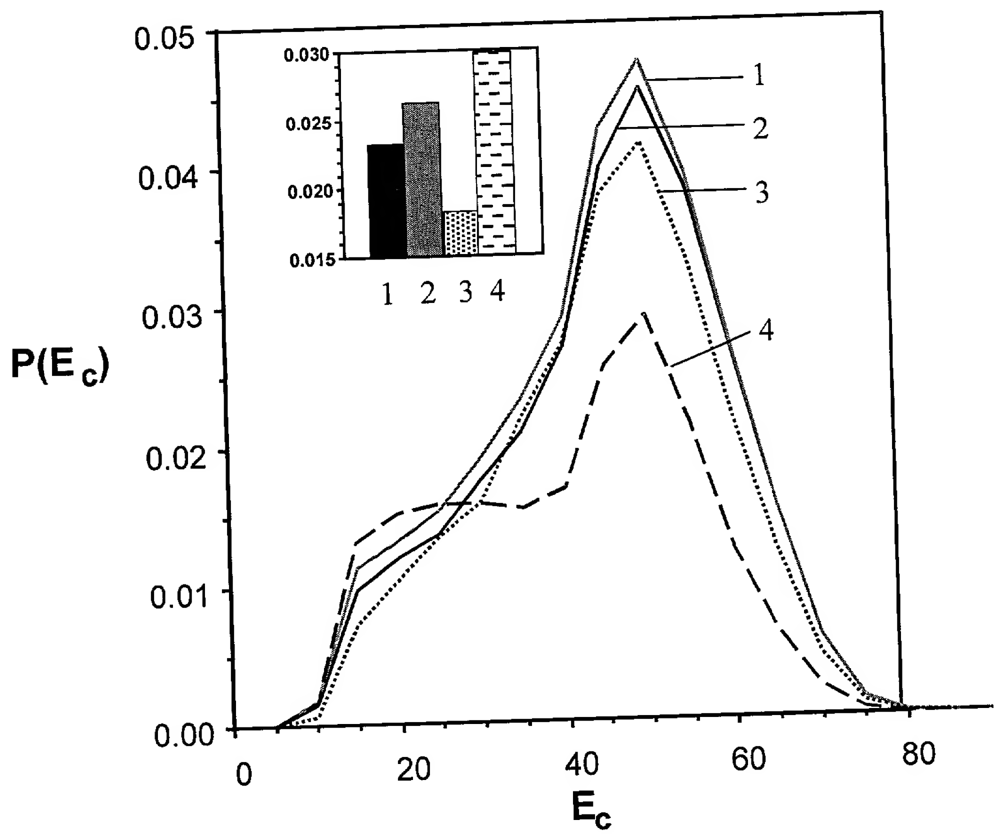

0.06
0.04
0.02
0.00
0
25
50
75
100
125
150
175
200
225
250
275
300
325
350
Residue Number

FIG. 4D



31
26
21
16
11
E_c
50
60
70
80
90
100
110
120
130
140
150
Residue Number
2

FIG. 5



0.05
0.04
0.03
0.02
0.01
0.00
$P(E_c)$
0
20
40
60
80
E_c
1
2
3
4
0.030
0.025
0.020
0.015
1 2 3 4

FIG. 6

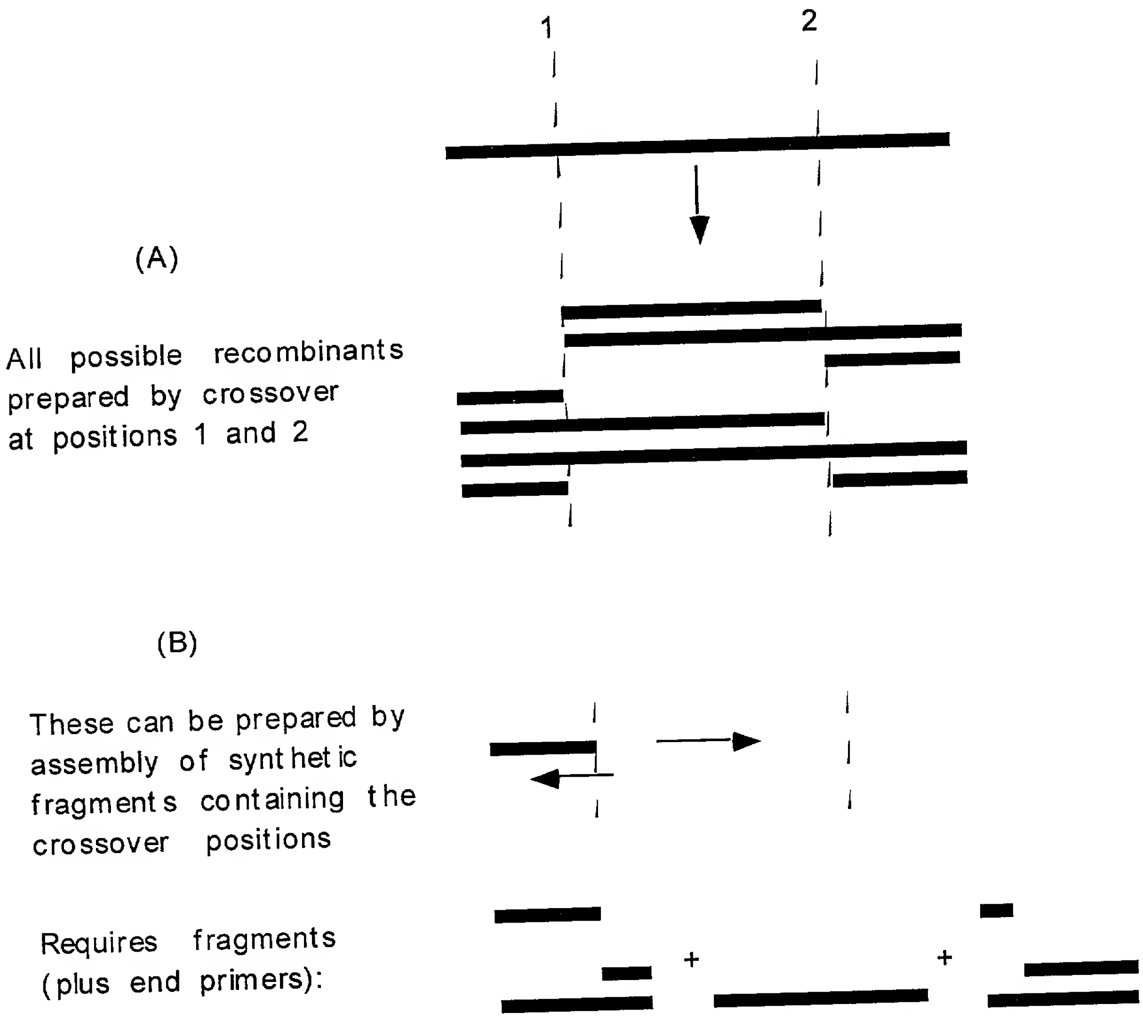

1
2
(A)
All possible recombinants prepared by crossover at positions 1 and 2
(B)
These can be prepared by assembly of synthetic fragments containing the crossover positions
Requires fragments (plus end primers):
+
+

FIG. 7

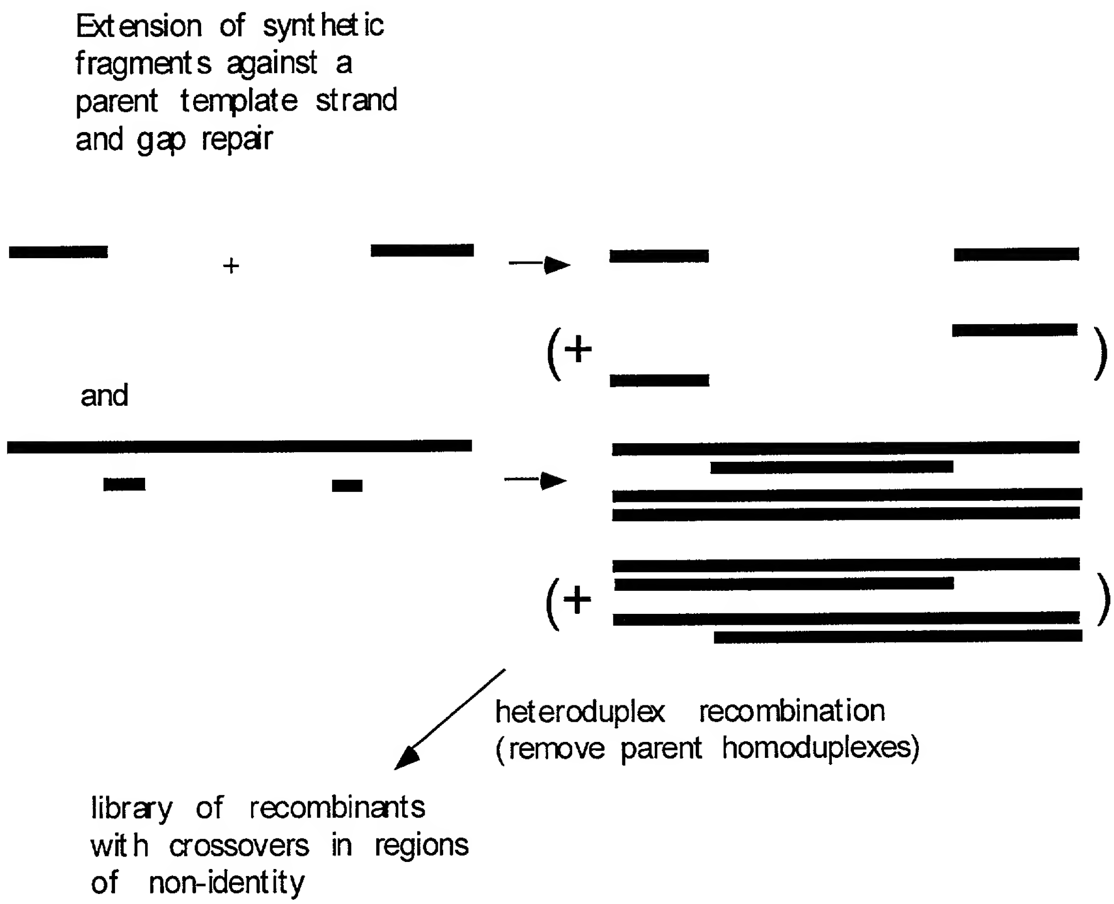
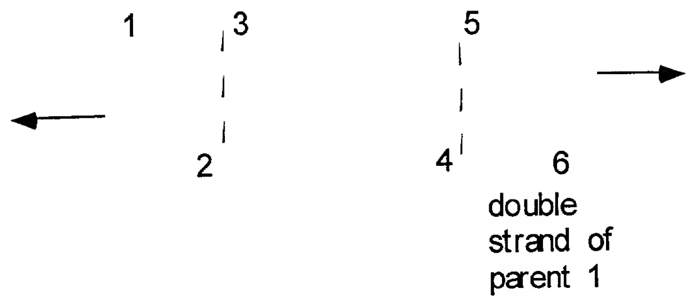

Extension of synthetic fragments against a parent template strand and gap repair
+
and
(+
)
(+
)
heteroduplex recombination (remove parent homoduplexes)
library of recombinants with crossovers in regions of non-identity

FIG. 8

(A)

Prepare the fragments by PCR with primers: perform reactions with primers 1+2, 3+4 and 5 +6,

and do same for other parent(s).



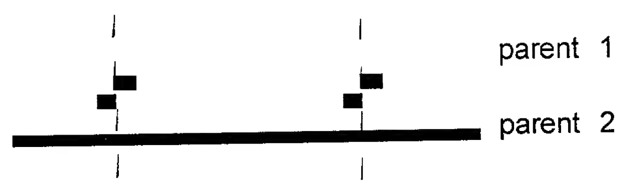
(B)

Reassemble fragments in a pool, by PCR with 1+ 6

FIG. 9

(A)

Prepare crossover primers designed to have crossovers at designated positions (2 primers for each position).



(B)

Fragment parent genes and PCR reassemble in the presence of the crossover primers to promote recombination at designated positions

FIG. 10

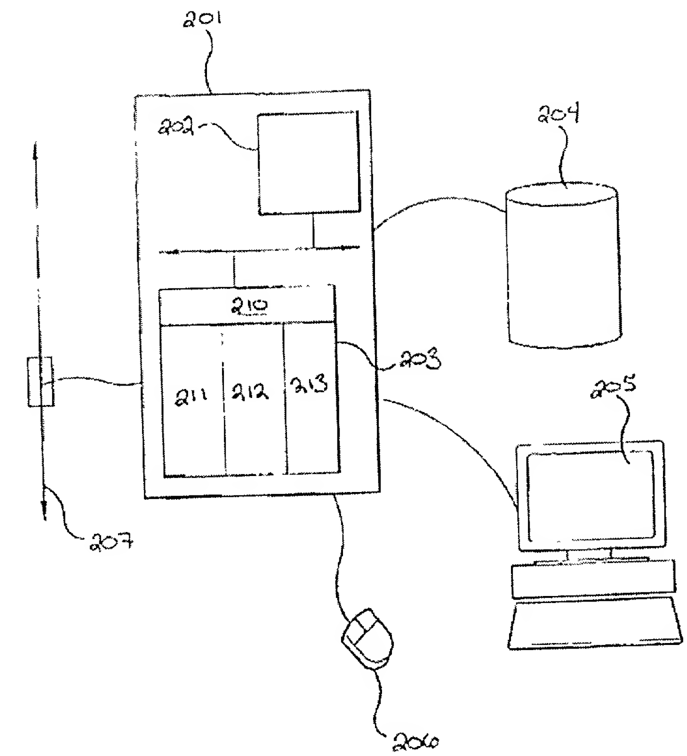


FIG. 11

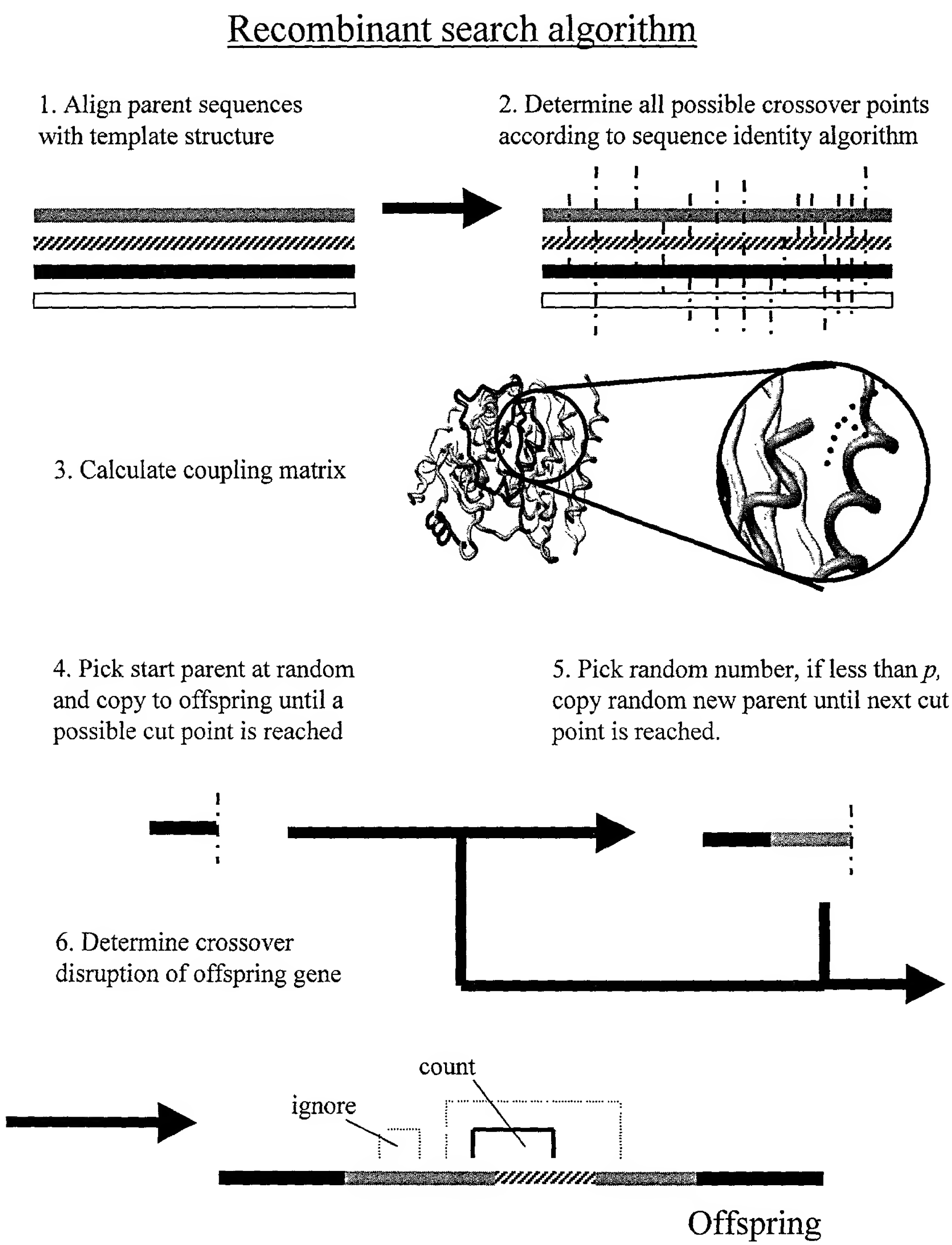
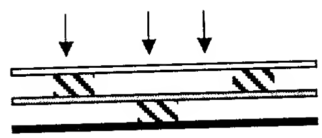
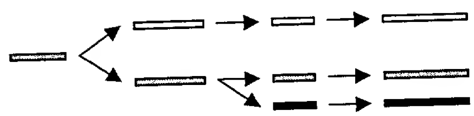

Recombinant search algorithm
1. Align parent sequences with template structure
2. Determine all possible crossover points according to sequence identity algorithm
3. Calculate coupling matrix
4. Pick start parent at random and copy to offspring until a possible cut point is reached
5. Pick random number, if less than p, copy random new parent until next cut point is reached.
6. Determine crossover disruption of offspring gene
count
ignore
Offspring

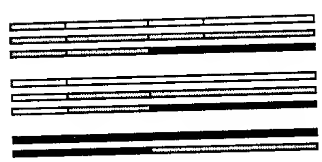
FIG. 12



(A)



(B)



(C)

FIG. 13

DIRECTED EVOLUTION ALGORITHM

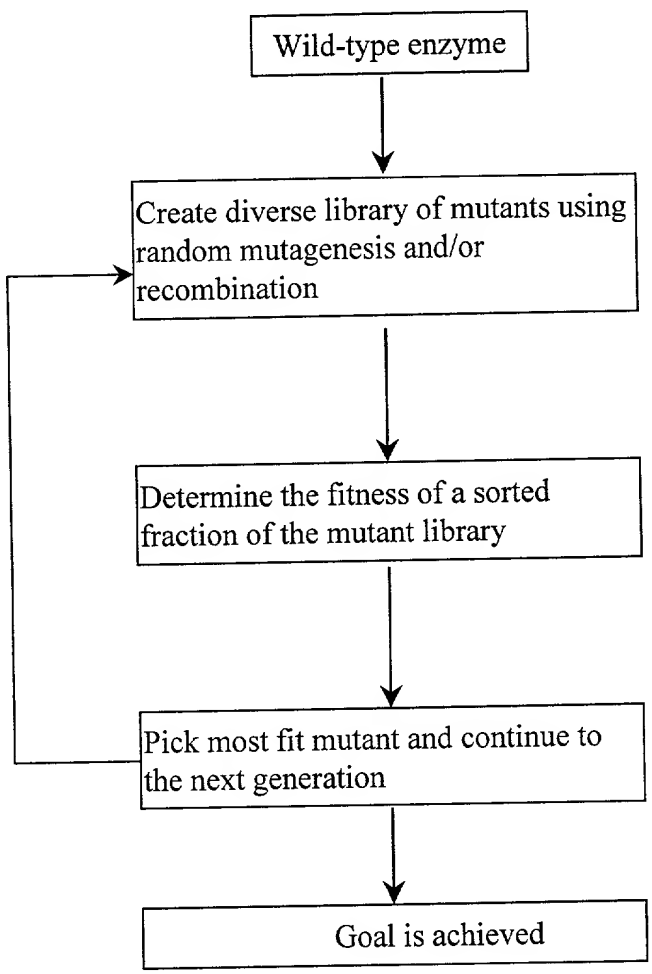


FIG. 14

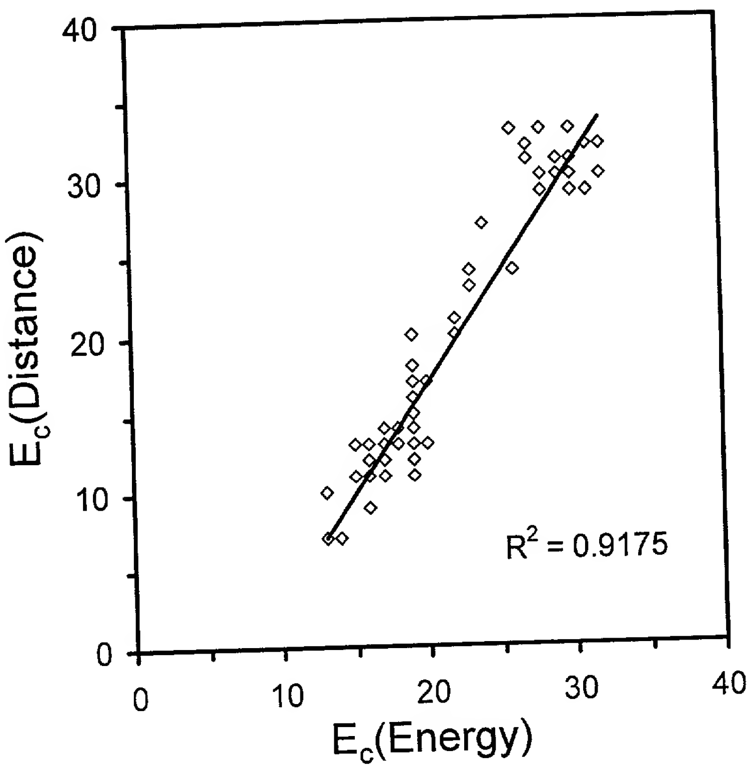
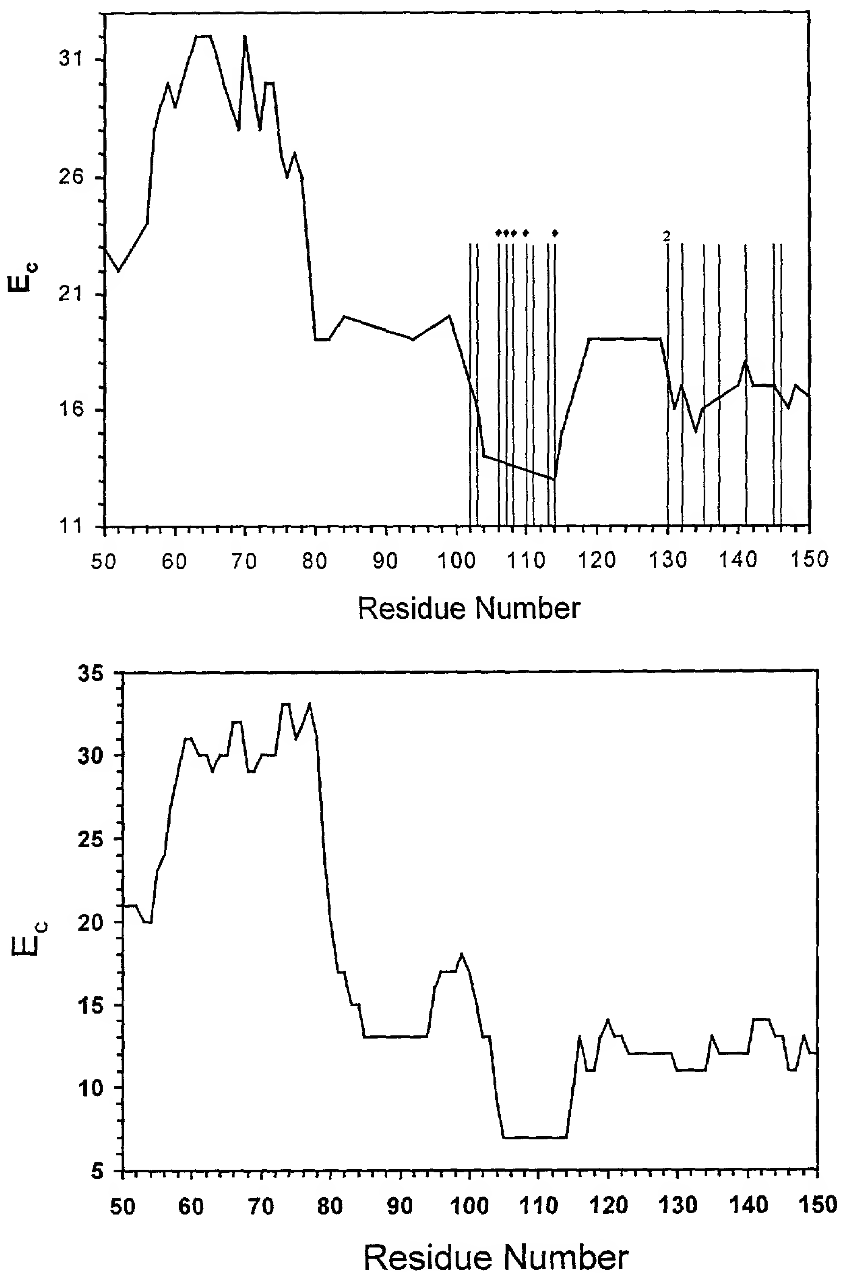


FIG. 15


31
26
21
16
11
E
c
50 60 70 80 90 100 110 120 130 140 150
Residue Number
2
35
30
25
20
15
10
5
E
c
50 60 70 80 90 100 110 120 130 140 150
Residue Number

(A)

(B)

FIG. 16

(A)

Experimental Data:

	wt	wt-insert	1	2
Tm(dC)	52	55.2	n.d.	54.3
Tm(dC)	49.5	53.3	44.5	52.5
t1/2	12.1	2586	-	87.5
t1/2	53	138	4	308

(B)

Calculations:

	All schema		Fragments		Z-score	
	av	stdev	1	2	1	2
Ec	19.260	4.090	10.770	8.124	-2.076	-2.723
Ec*	0.006	0.002	0.014	0.005	4.838	-0.857

FIG. 17

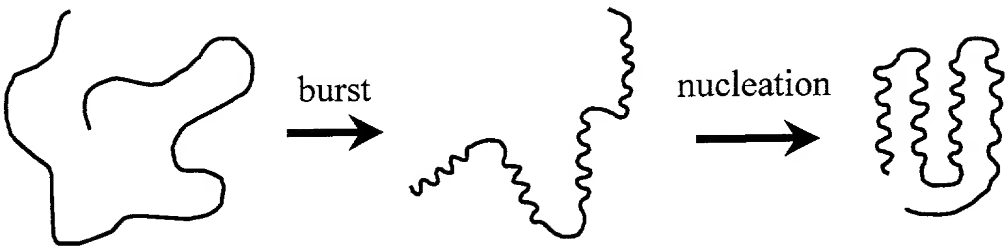


FIG. 18

The contact map shows residues that are distant (black) and residues that are close (white). If a given segment, , folds an above average number of residues into a given sphere size, then it is compact.

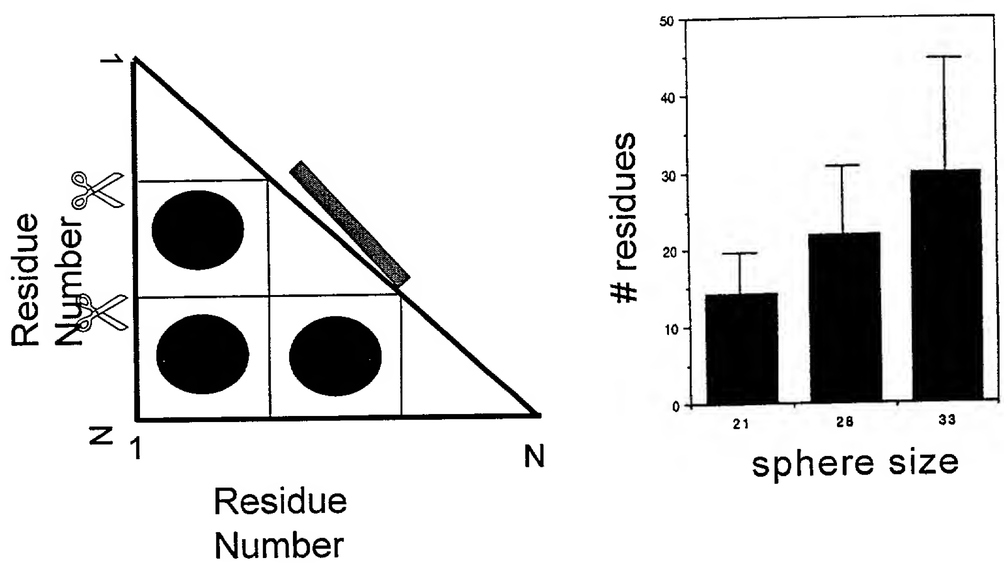
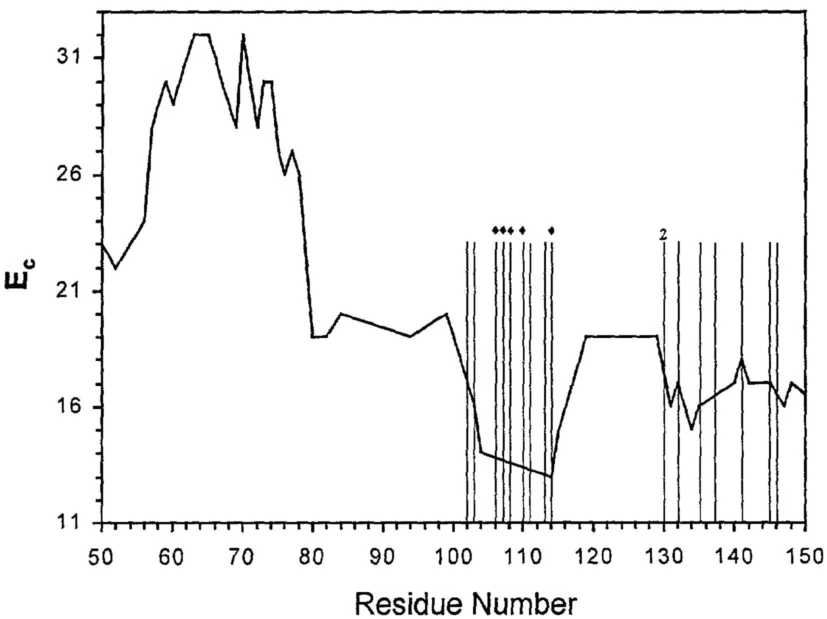
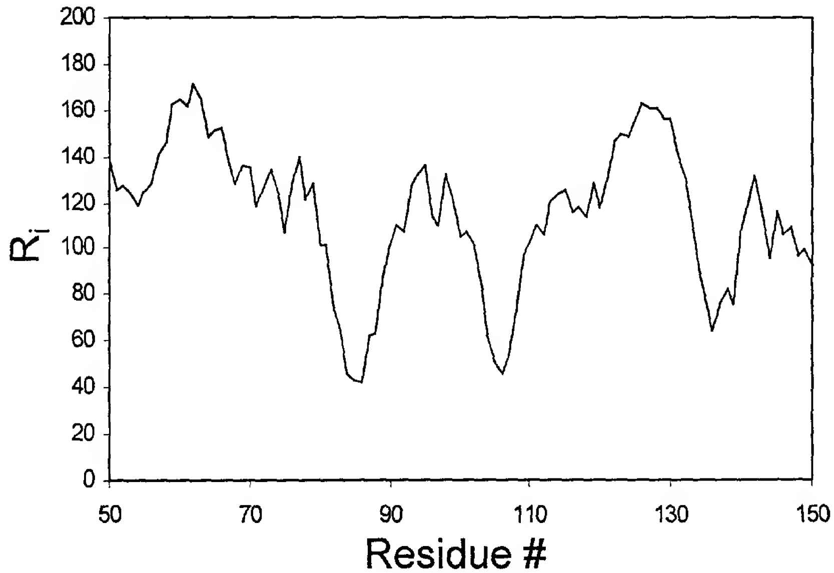


FIG. 19


31
26
21
16
11
E_c
50 60 70 80 90 100 110 120 130 140 150
Residue Number
2

(A)


200
180
160
140
120
100
80
60
40
20
0
R_i
50 70 90 110 130 150
Residue #

(B)

FIG. 20

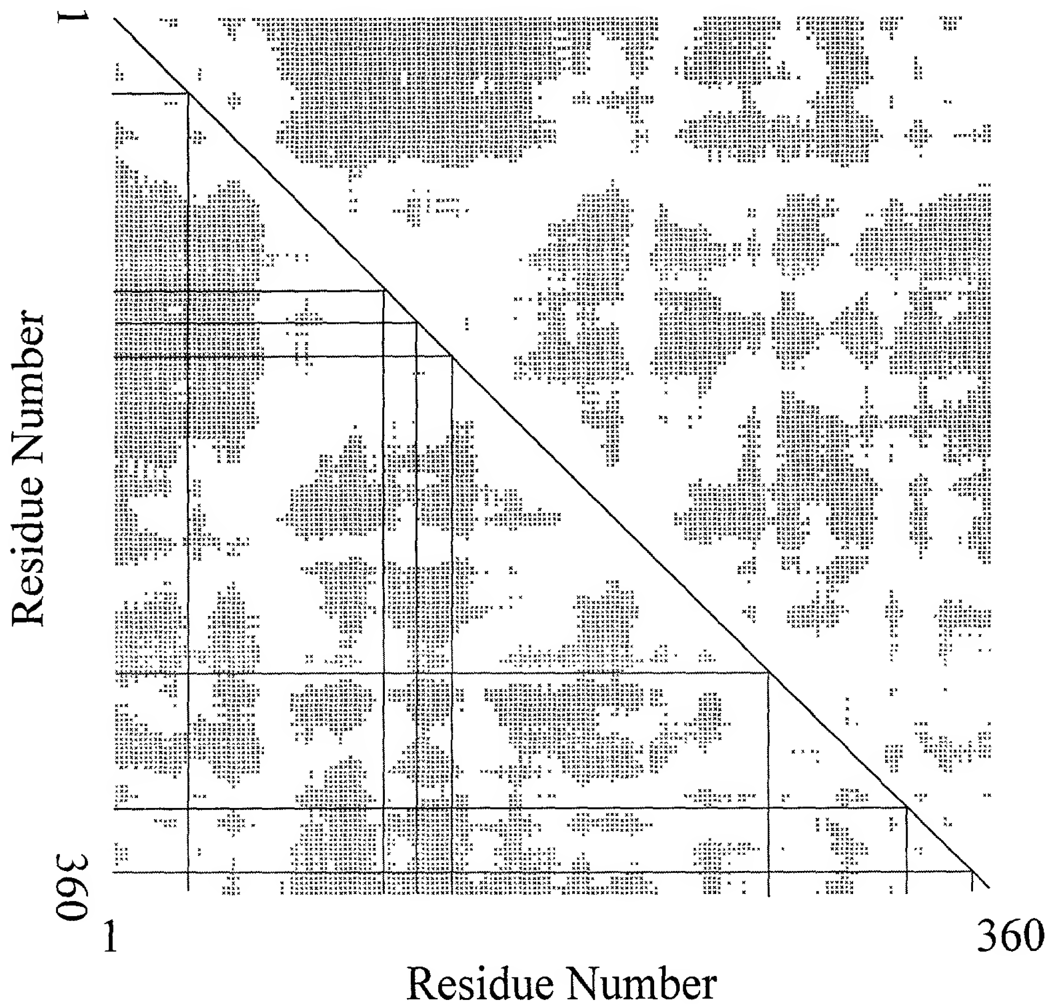


FIG. 21

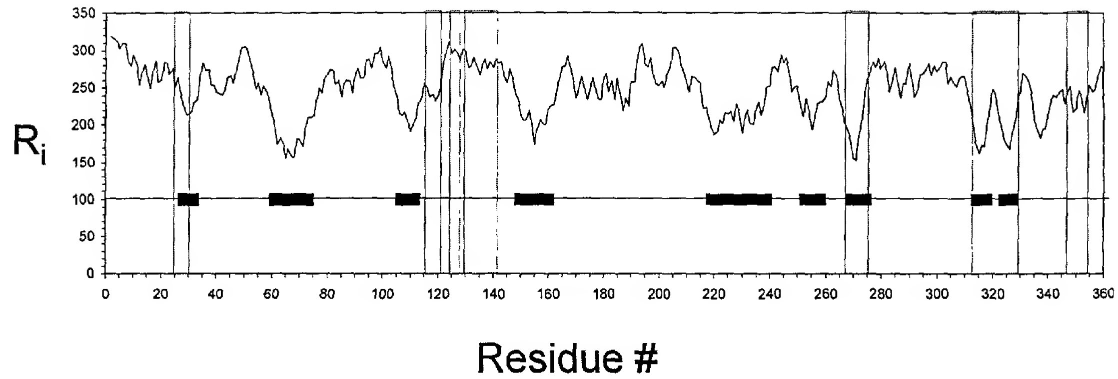
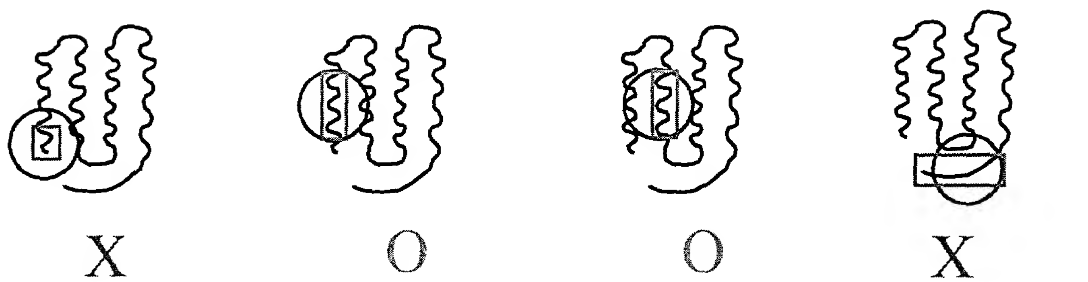


FIG. 22



Is rejected because the fragment size is less than 15

Is accepted because the fragment is > 15 residues and fits into the sphere

Is accepted because the fragment is > 15 residues and fits into the sphere

Is rejected, because the fragment does not fit into the sphere

(1) Pick a sphere size (21 angstroms, like Go-Gilbert) and a disruption threshold; (2) Scan protein using segments <u>at least</u> the average number of residues for that sphere size <u>or greater</u> (e.g., >15 for 21 angstrom sphere); 3) Check the disruption of all the compact fragments identified in step 2. If the fragment has a disruption above a threshold value, keep it; otherwise, throw it out; 4) If the compact unit is disruptive, increment the schema disruption measure for all of the residues in the fragment by one. This indicates that crossovers within the fragment are disfavored.

FIG. 23

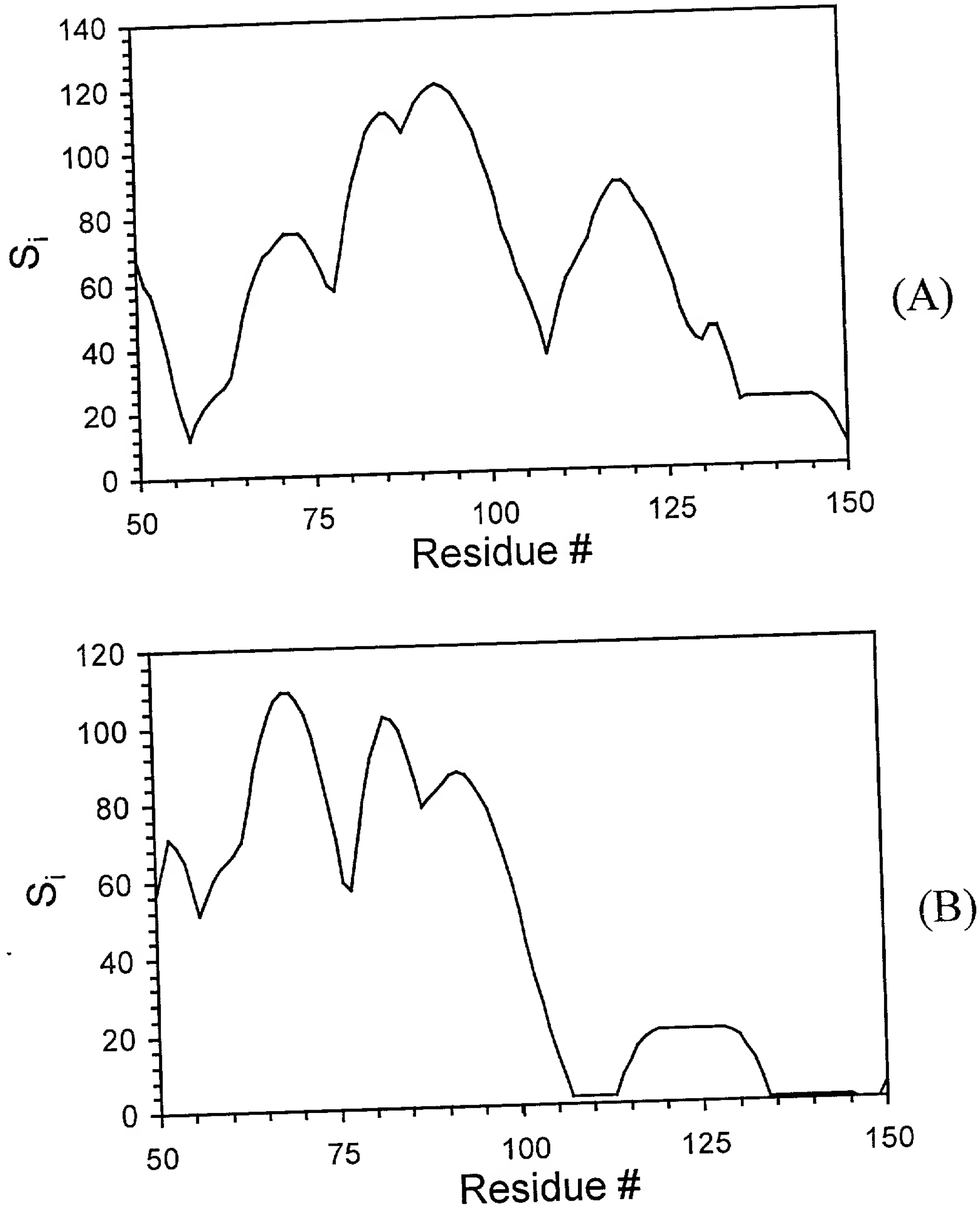

140
120
100
80
60
40
20
0
S_i
50
75
100
125
150
Residue #
(A)
120
100
80
60
40
20
0
S_i
50
75
100
125
150
Residue #
(B)

FIG. 24

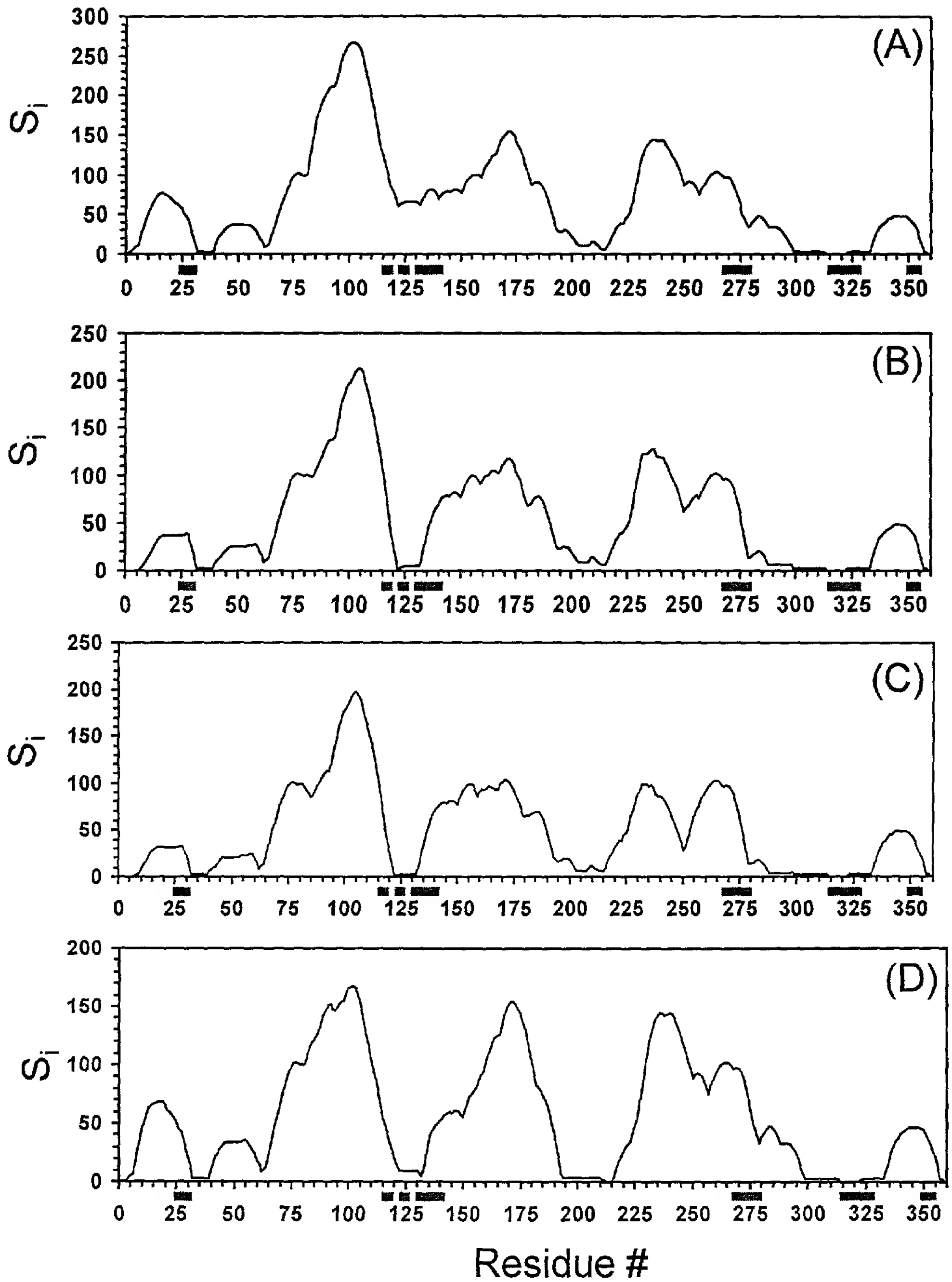


FIG. 25

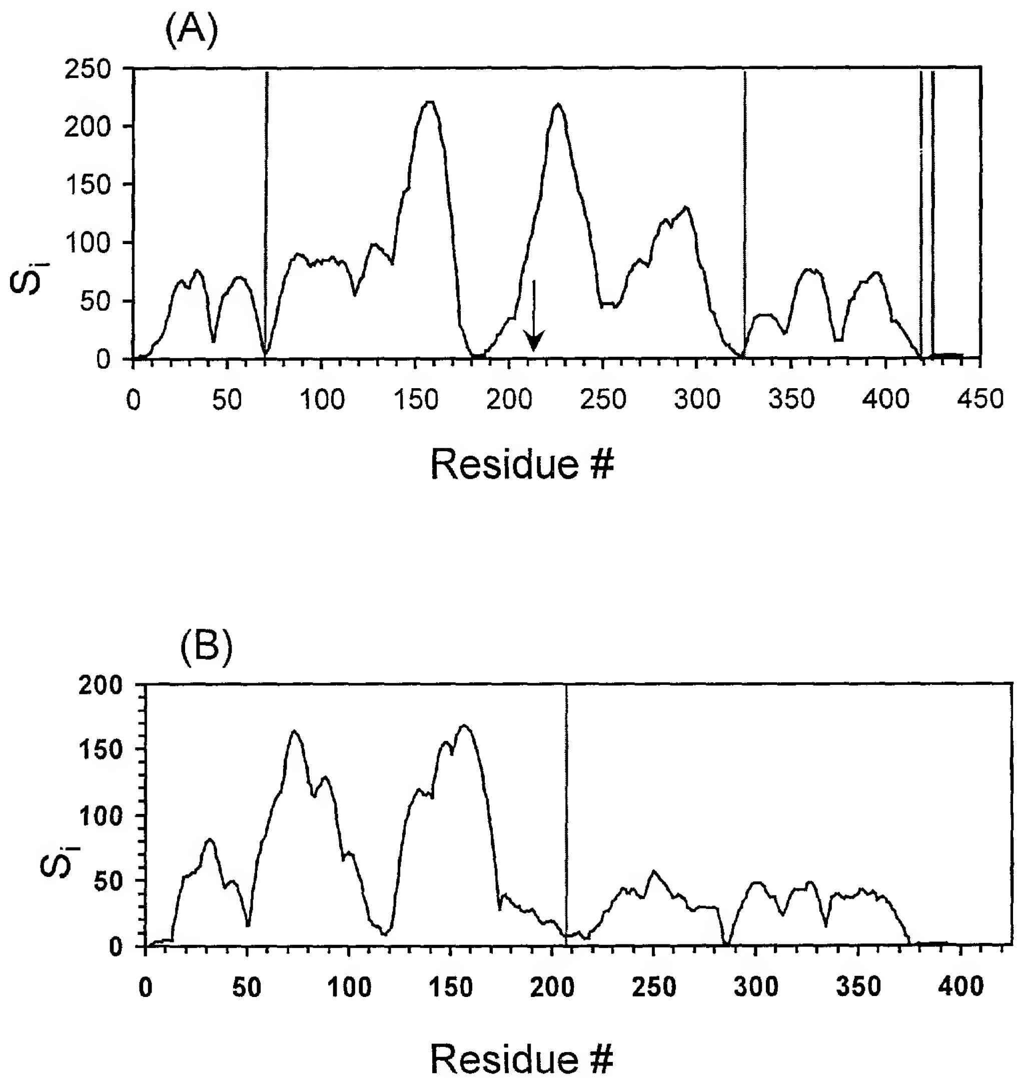

(A)
250
200
150
100
50
0
S_i
0
50
100
150
200
250
300
350
400
450
Residue #
(B)
200
150
100
50
0
S_i
0
50
100
150
200
250
300
350
400
Residue #

FIG. 26

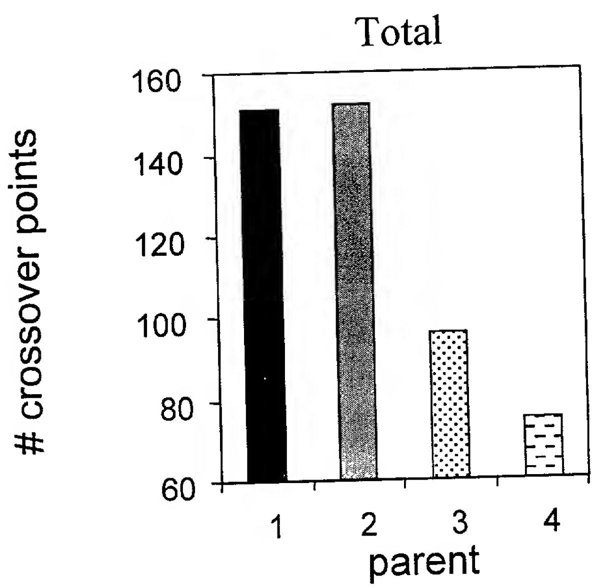


FIG. 27A

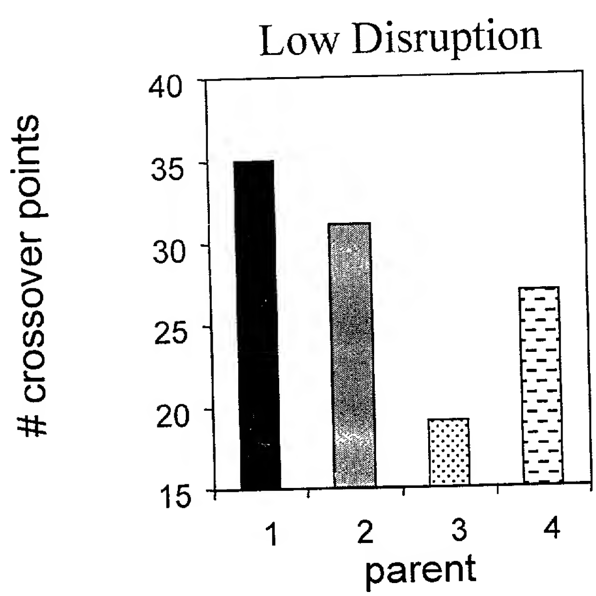


FIG. 27B